Latent Variable Example

This example shows the relationship between the X loadings from the SIMPLS algorithm and the latent variables.

The concentration matrix C_{train} contains the percent by weight of each of 20 solutions in the training data. The columns correspond to benzene, polystyrene, and gasoline. In this example, you will compute the latent variables for each of the constituents individually by using one column of C_{train} at a time. Computing partial least squares (PLS) on one constituent at a time is called PLS1. Computing PLS on more than one constituent at a time is called PLS2.

$$
C_{train} =
\begin{array}{c|ccc|c}
 & \text{benzene} & \text{polystyrene} & \text{gasoline} & \\
\hline
 & 0 & 0 & 100.0000 & 1 \\
 & 5.1309 & 0 & 94.8691 & 2 \\
 & 10.0660 & 0 & 89.9300 & 3 \\
 & 20.1799 & 0 & 79.8201 & 4 \\
 & 40.0120 & 0 & 59.9878 & 5 \\
 & 59.9972 & 0 & 40.0028 & 6 \\
 & 79.8412 & 0 & 20.1588 & 7 \\
 & 89.8273 & 0 & 10.1727 & 8 \\
 & 100.0000 & 0 & 0 & 9 \\
 & 90.0264 & 9.9736 & 0 & 10 \\
 & 80.1375 & 19.8625 & 0 & 11 \\
 & 64.9950 & 35.0005 & 0 & 12 \\
 & 21.0228 & 45.9197 & 33.0575 & 13 \\
 & 49.9507 & 5.0599 & 44.9895 & 14 \\
 & 40.0182 & 20.0385 & 39.9433 & 15 \\
 & 40.0154 & 10.0036 & 49.9810 & 16 \\
 & 30.0059 & 10.0282 & 59.9659 & 17 \\
 & 40.0340 & 39.9670 & 19.9990 & 18 \\
 & 49.9393 & 3.3748 & 46.6859 & 19 \\
 & 46.6501 & 13.4658 & 39.8840 & 20 \\
\end{array}
$$

Load the napalm data set.

```
clearvars
load pnnl_napalm_data
```

n is the number of constituents.

```
n = size(C_train,2)
```

n = 3

PLS1 Without Mean Centering

Compute the X loadings using the SIMPLS algorithm without mean centering the data. Computing partial least squares (PLS) on one constituent at a time is called PLS1.

Set the number of latent variables.

```
nLatentVariables = 10;
```

Set up for plotting.

```
clf
color = pnnl_colorOrder(n);
```

For each of the constituents in the columns of C_{train}, compute the latent variables without mean-centered data using the SIMPLS algorithm. We're plotting in reverse order so benzene is the top plot.

```
for k = n:-1:1
    X = A_train;
    Y = C_train(:,k);
    [X_loadings, Y_loadings, X_scores, Y_scores, Weights] = pnnl_simpls(X, Y, nLatentVa
```

The latent variables (LV) are the rectified and normalized X loadings from the SIMPLS algorithm. The X loadings are column-oriented, so transpose them so the latent variables will be row-oriented.
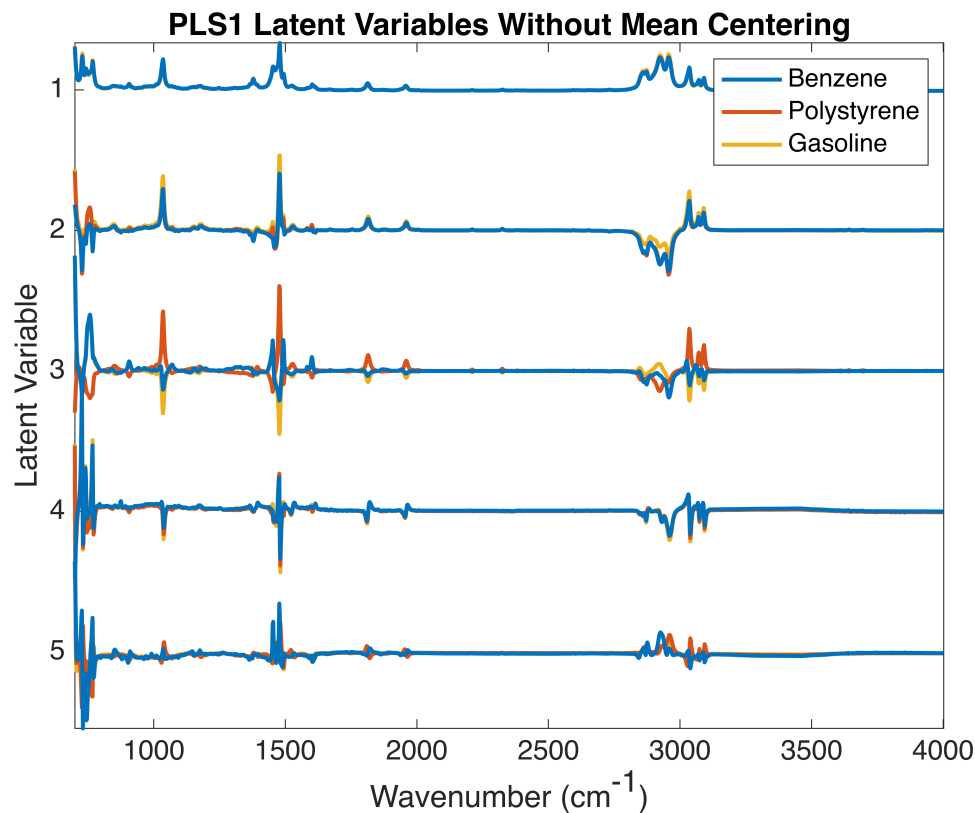
```
    LV = pnnl_normalize_rows(pnnl_rectify_rows(X_loadings'));
```

Plot the first five latent variables.

```
    h{k} = pnnl_strip_plot(Wavenumbers,LV(1:5,:),'PLS1 Latent Variables Without Mean C
    hold on
end
```

Turn on the legend for each constituent.

```
legend([h{1}(1) h{2}(1) h{3}(1)])
```

2

PLS1 Latent Variables Without Mean Centering

PLS1 With Mean Centering

Compute the X loadings using the SIMPLS algorithm with mean centering the data. Computing partial least squares (PLS) on one constituent at a time is called PLS1.

Set the number of latent variables.

```
nLatentVariables = 10;
```

Set up for plotting.

```
clf
color = pnnl_colorOrder(n);
```

For each of the constituents in the columns of C_{train}, compute the latent variables without mean-centered data using the SIMPLS algorithm. We're plotting in reverse order so benzene is the top plot.

```
for k = n:-1:1
    X = A_train;
    Y = C_train(:,k);
```

Mean-center the data.

```
    X0 = X - mean(X,1);
    Y0 = Y - mean(Y,1);
```

Compute the loadings using the SIMPLS algorithm.

```
[X_loadings, Y_loadings, X_scores, Y_scores, Weights] = pnnl_simpls(X0, Y0, nLaten
```

The latent variables (LV) are the rectified and normalized X loadings from the SIMPLS algorithm. The X loadings are column-oriented, so transpose them so the latent variables will be row-oriented.
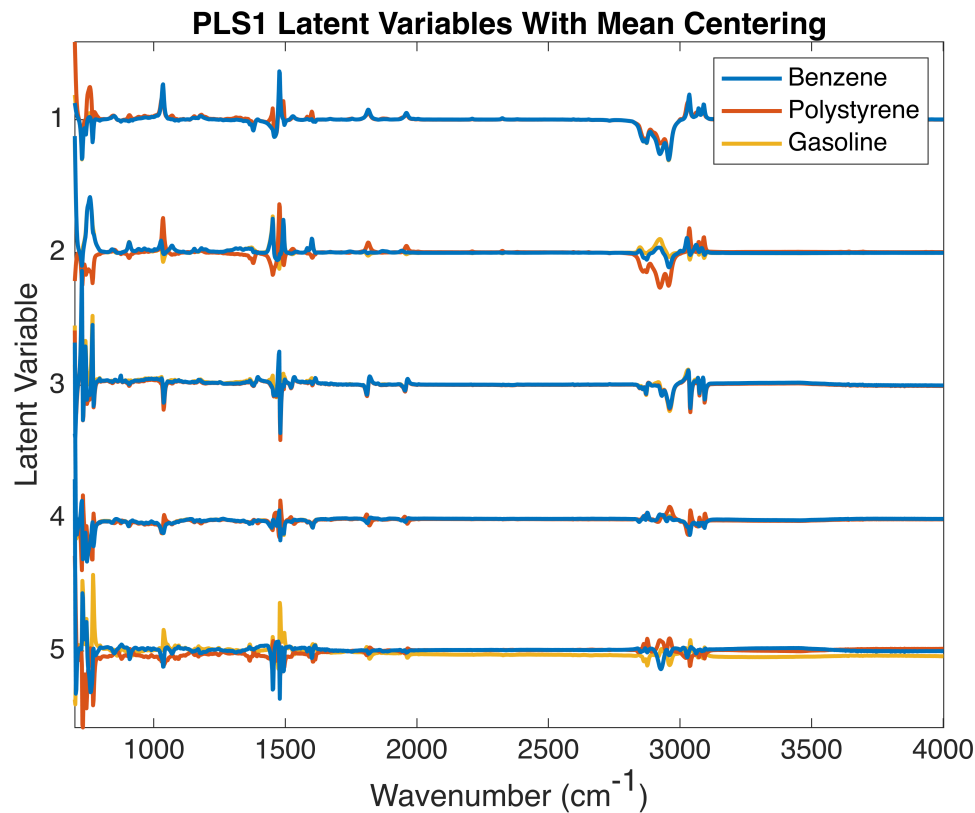
```
LV = pnnl_normalize_rows(pnnl_rectify_rows(X_loadings'));
```

Plot the first five latent variables.

```
h{k} = pnnl_strip_plot(Wavenumbers,LV(1:5,:),'PLS1 Latent Variables With Mean Cente
hold on
end
```

Turn on the legend for each constituent.

```
legend([h{1}(1) h{2}(1) h{3}(1)])
```


PLS2 Without Mean Centering

Compute the X loadings using the SIMPLS algorithm without mean centering the data. Computing partial least squares (PLS) on more than one constituent at a time is called PLS2.

Set the number of latent variables.

```
nLatentVariables = 10;
```

For all of the constituents in the columns of C_{train}, compute the latent variables without mean-centered data using the SIMPLS algorithm.
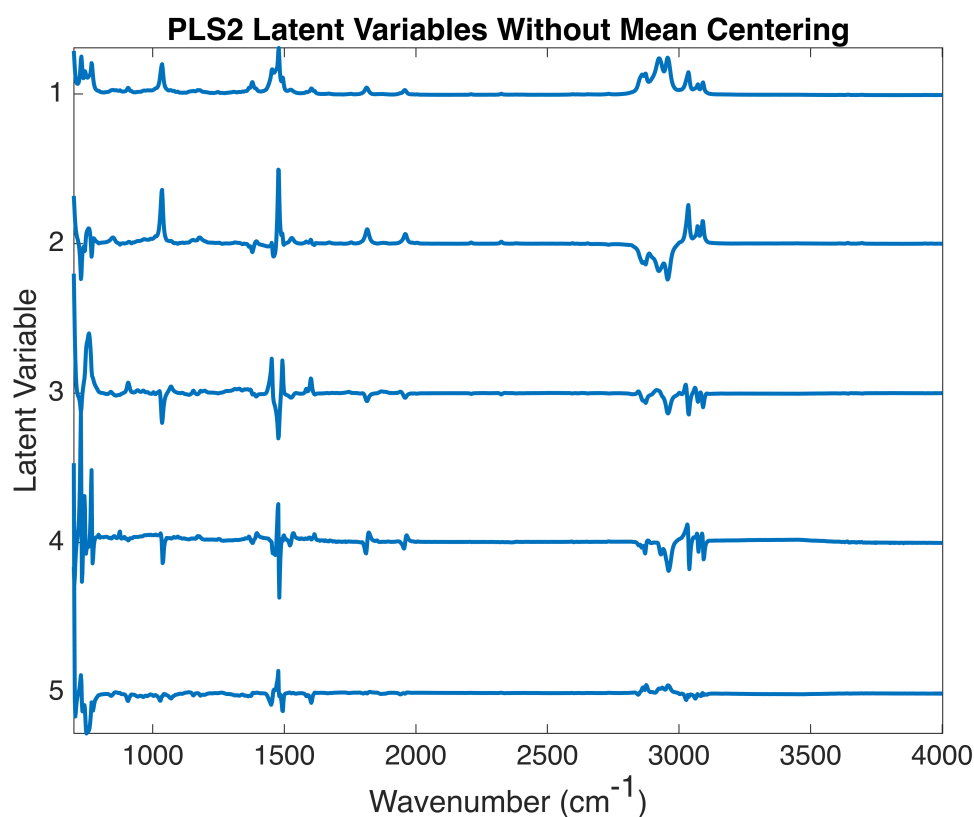
```
X = A_train;
Y = C_train;
[X_loadings, Y_loadings, X_scores, Y_scores, Weights] = pnnl_simpls(X, Y, nLatentVarial
```

The latent variables (LV) are the rectified and normalized X loadings from the SIMPLS algorithm. The X loadings are column-oriented, so transpose them so the latent variables will be row-oriented.

```
LV = pnnl_normalize_rows(pnnl_rectify_rows(X_loadings'));
```

Plot the first five latent variables.

```
clf
pnnl_strip_plot(Wavenumbers,LV(1:5,:),'PLS2 Latent Variables Without Mean Centering',Wa
```


PLS2 With Mean Centering

Compute the X loadings using the SIMPLS algorithm without mean centering the data. Computing partial least squares (PLS) on more than one constituent at a time is called PLS2.

Set the number of latent variables.

```
nLatentVariables = 10;
```

For all of the constituents in the columns of C_{train}, compute the latent variables without mean-centered data using the SIMPLS algorithm.

```
X = A_train;
Y = C_train;
```

Mean-center the data.

```
X0 = X - mean(X,1);
Y0 = Y - mean(Y,1);
```

Compute the loadings using the SIMPLS algorithm.

```
[X_loadings, Y_loadings, X_scores, Y_scores, Weights] = pnnl_simpls(X0, Y0, nLatentVar
```

The latent variables (LV) are the rectified and normalized X loadings from the SIMPLS algorithm. The X loadings are column-oriented, so transpose them so the latent variables will be row-oriented.

```
LV = pnnl_normalize_rows(pnnl_rectify_rows(X_loadings'));
```

Plot the first five latent variables.

```
clf
pnnl_strip_plot(Wavenumbers,LV(1:5,:),'PLS2 Latent Variables With Mean Centering',Waven
```


PLS2 Latent Variables With Mean Centering
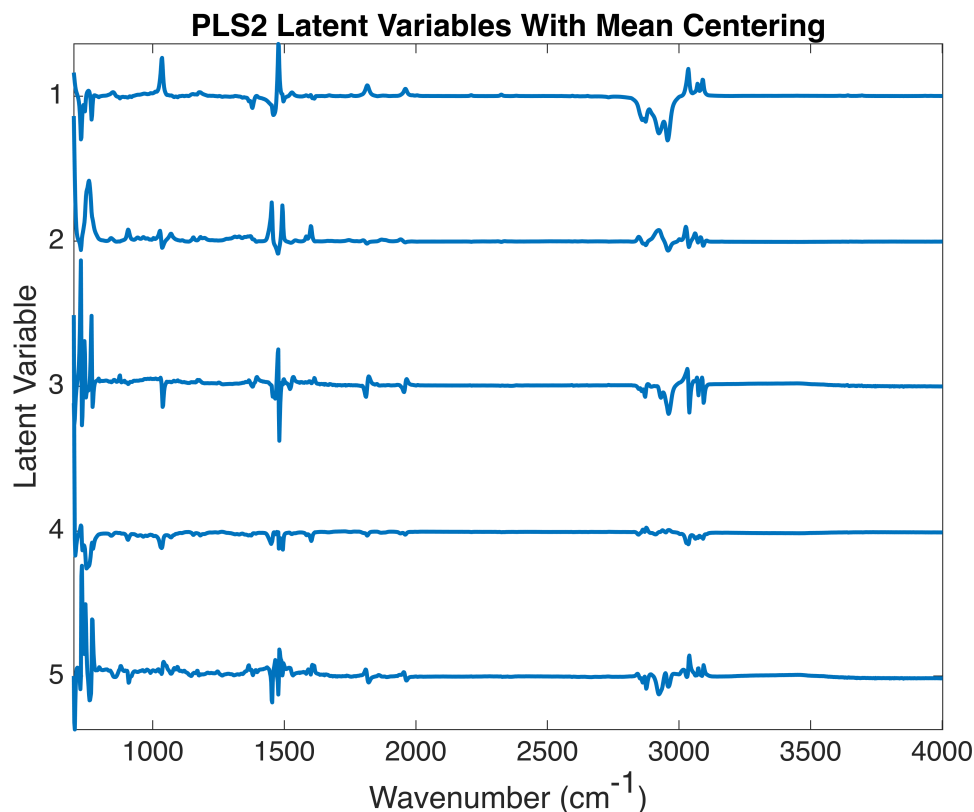
Suppress warning messages.

```
%#ok<*ASGLU>
```

Disclaimer

This material was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the United States Department of Energy, nor Battelle, nor any of their employees, nor any jurisdiction or organization that has cooperated in the development of these materials, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness or any information, apparatus, product, software, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC05-76RL01830